MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2017/2018

BMD3024 – MULTIVARIATE DATA ANALYSIS

(All sections / Groups)

11 OCTOBER 2017
2.30 p.m. - 4.30 p.m.
(2 Hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 7 pages (including cover page) with 4 structured questions.
2. Answer **ALL** questions.
3. Marks are shown at the end of each question.
4. Write all your answers in the answer booklet provided.

Question One (25 marks)

(a) Describe the **TWO (2)** broad types of multivariate methods.

(4 marks)

(b) Explain **THREE (3)** methods and tests used to examine the assumptions of:-

(i) Normality

(3 marks)

(ii) Homoscedasticity.

(3 marks)

(c) **Table 1** shows the valid data, missing data, missing data percentage, mean, and standard deviation values for V1, V2, V7, and V13.

Assume that the missing data below are missing completely at random (MCAR).

Table 1

	Day1	**Day2**	**Day3**
Valid	810	546	687
Missing	90	264	59
Missing (%)	11.11%	48.35%	8.6%

Apply the 'Rules of Thumb' on Imputation of Missing Data and relate the preferred data imputation techniques for the following variables:

(i) Day1.

(3 marks)

(ii) Day2.

(3 marks)

(iii) Day3.

(3 marks)

(d) Describe the sample size considerations that must be made when conducting multiple regression.

(6 marks)

Question Two (25 marks)

(a) Explain the **FOUR (4)** criteria for the number of factors to extract when conducting Factor Analysis.

(8 marks)

Continued...

(b) A study was carried out to explore the relationship between *Aggression* and several potential predicting factors in 666 children that had an older sibling. Variables measured were *Parenting Style* (high score = bad parenting practices), *Computer Games* (high score = more time spent playing computer games), *Television* (high score = more time spent watching television), *Diet* (high score = the child has a good diet low in E-numbers), and *Sibling Aggression* (high score = more aggression seen in their older sibling). Past research indicated that parenting style and sibling aggression were good predictors of the level of aggression in the younger child.

Table 2: Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	0.287[a]	0.083	0.076	0.30709	1.913

Table 3: ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	**Regression**	5.602	5	1.120	11.882	0.000[b]
	Residual	62.240	660	0.094		
	Total	67.842	665			

a. Dependent Variable: Aggression
b. Predictors: (Constant), Parenting Style, Computer Games, Television, Diet, Sibling Aggression

Table 4: Coefficients[a]

Model	Unstandardized Coefficients		Standard-ized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tole-rance	VIF
(Constant)	-0.005	0.012		-0.416	0.677	-	-
Parenting Style	0.057	0.015	0.177	3.891	0.000	0.669	1.494
Computer Games.	0.142	0.037	0.152	3.851	0.000	0.891	1.123
Television	0.033	0.046	0.032	0.715	0.475	0.697	1.436
Diet	-0.109	0.038	-0.115	-2.864	0.004	0.862	1.160
Sibling Aggression	0.082	0.039	0.084	2.106	0.036	0.883	1.133

a. Dependent Variable: Aggression

Continued...

(i) Refer to **Table 2**. Why is the adjusted R^2 lower than the unadjusted R^2?
(4 marks)

(ii) Refer to **Table 3**. Is the *F*-statistic significant? What does this indicate?
(4 marks)

(iii) According to **Table 4**, which are the variables that are significant at $p<0.05$ in the prediction? Explain your answer.
(3 marks)

(iv) According to **Table 4**, is there a multicollinearity issue in the multiple regression model? Explain your answer.
(3 marks)

(v) 'Multicollinearity generally viewed as harmful'. Discuss.
(3 marks)

Question Three (25 marks)

(a) Researchers were interested in the effects of cognitive behaviour therapy on obsessive compulsive disorder (OCD). They compared a group of OCD sufferers after cognitive behaviour therapy (*CBT*) and after behaviour therapy (*BT*) with a group of OCD sufferers who are still awaiting treatment (a no-treatment condition, *NT*). There are two dependent measures: the occurrence of obsession-related behaviours (*Actions*) and the occurrence of obsession-related cognitions (*Thoughts*). These dependent variables were measured on a single day and so represent the number of obsession-related behaviours/thoughts in a normal day.

The research's outputs are found in **Table 5** to **Table 7**. Use these tables to answer questions (a) (i), (ii), and (ii).

Table 5: Box's Test of Equality of Covariance Matrices[a]

Box's M	9.959
F	1.482
df1	6
df2	18168.923
Sig.	0.180

Continued...

Table 6: Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	0.983	745.230[b]	2.000	26.000	0.000
	Wilks' Lambda	0.017	745.230[b]	2.000	26.000	0.000
	Hotelling's Trace	57.325	745.230[b]	2.000	26.000	0.000
	Roy's Largest Root	57.325	745.230[b]	2.000	26.000	0.000
Group	Pillai's Trace	0.318	2.557	4.000	54.000	0.049
	Wilks' Lambda	0.699	2.555[b]	4.000	52.000	0.050
	Hotelling's Trace	0.407	2.546	4.000	50.000	0.051
	Roy's Largest Root	0.335	4.520[c]	2.000	27.000	0.020

a. Design: Intercept + Group
b. Exact statistic
c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Table 7: Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Actions	10.467[a]	2	5.233	2.771	0.080
	Thoughts	19.467[b]	2	9.733	2.154	0.136
Intercept	Actions	616.533	1	616.533	326.400	0.000
	Thoughts	6336.533	1	6336.533	1402.348	0.000
Group	Actions	10.467	2	5.233	2.771	0.080
	Thoughts	19.467	2	9.733	2.154	0.136
Error	Actions	51.000	27	1.889		
	Thoughts	122.000	27	4.519		
Total	Actions	678.000	30	-	-	-
	Thoughts	6478.000	30	-	-	-
Corrected Total	Actions	61.467	29	-	-	-
	Thoughts	141.467	29	-	-	-

(i) Interpret the results in **Table 5**. Was the assumption of homogeneity of covariance violated? (2 marks)

(ii) Based on the output in **Table 6**, is the variable *Group* a significant predictor for the dependent variate, using Pillai's Trace statistic? (4 marks)

(iii) Based on the output in **Table 7**, examine and report the results of the univariate effects of *Group* on each of the dependent variables at 5% level of significance. (6 marks)

Continued...

(b) Dr Sian Williams (University of Brighton) devised a questionnaire to measure organizational ability. She predicted five factors to do with organizational ability, and these dimensions are theoretically independent. Dr Sian Williams' questionnaire contains 28 items using a 7-point Likert scale (1 = strongly disagree, 4 = neither, 7 = strongly agree). She gave it to 239 people.

Table 8 is the **unrotated** component matrix of the Principal Component Analysis that was run on the data from Dr. Sian Williams' research.

Table 8: Unrotated Component Matrix[a]

	Component				
	1	**2**	**3**	**4**	**5**
org1	.684	-.341			
org2	.373	-.543		.303	.312
org3	.584			.313	
org4	.600			.452	
org6	.446		.524		
org7		-.501			.453
org9	.528	.394			
org10	.803				
org11	.723	-.300			
org12	.502				
org13	.675				
org14			.519		
org16	.673				
org17	.614			-.517	
org18	.559				-.390
org19	.650			-.497	
org20	.768				
org21	.421	-.523			
org22		.620			.325
org23	.456	.382			
org24	.674				
org25	.791				
org26	.432		.518		
org27	.614			.345	-.334
org28	.501	.444			

a. 5 components extracted.

(i) Examine the distribution of variables amongst the six (6) factors.

(4 marks)

(ii) Based on the results in (b(i)), should the variables be rotated? Explain.

(3 marks)

Continued...

(c) Explain the **THREE (3)** steps to perform a multidimensional scaling analysis.
(6 marks)

Question Four (25 marks)

A personality test containing 28 questions (items) was distributed to 2,800 respondents. Out of all the questions, the first 25 questions measure five (5) personality traits or **factors**: *Agreeableness* (A), *Conscientiousness* (C), *Extraversion* (E), *Neuroticism* (N), and *Openness* (O). The variables are labelled A1-A5, C1-C5, E1-E5, N1-N5, and O1-O5.

Each factor is measured using five (5) questions or variable. The variables are labelled A1-A5, C1-C5, E1-E5, N1-N5, and O1-O5.

(a) Construct **one (1)** path diagram that shows the relationship between 25 variables and 5 factors. **Label** all variables, factors, and error terms of the variables.
(23 marks)

(b) In the same path diagram drawn in (a), draw and label the following hypothesized structural relationships:

H_1: *Agreeableness (A)* is correlated to *Openness (O)*.

H_2: *Conscientiousness (C)* is correlated to *Neuroticism (N)*.
(2 marks)

End of Page